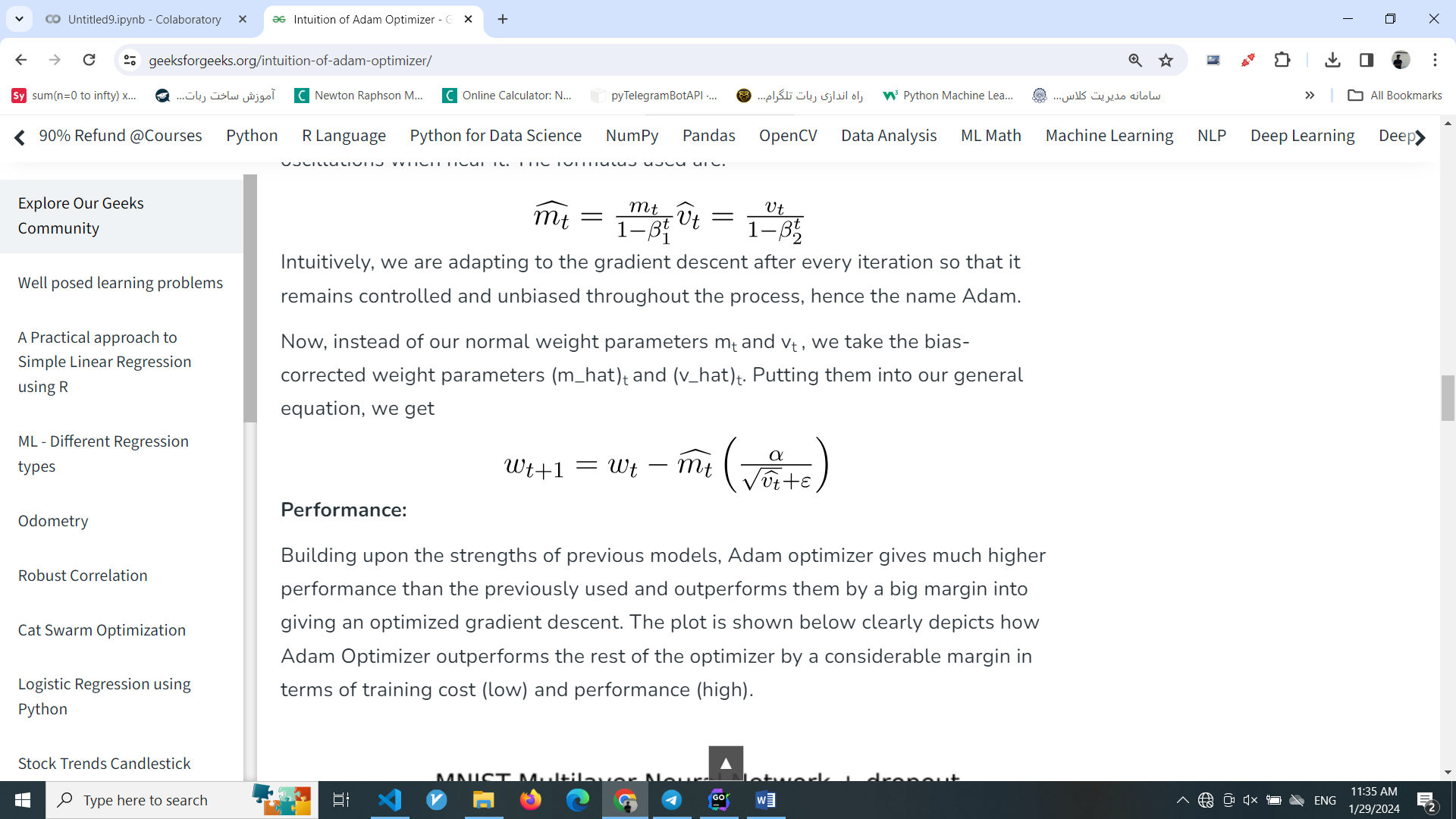
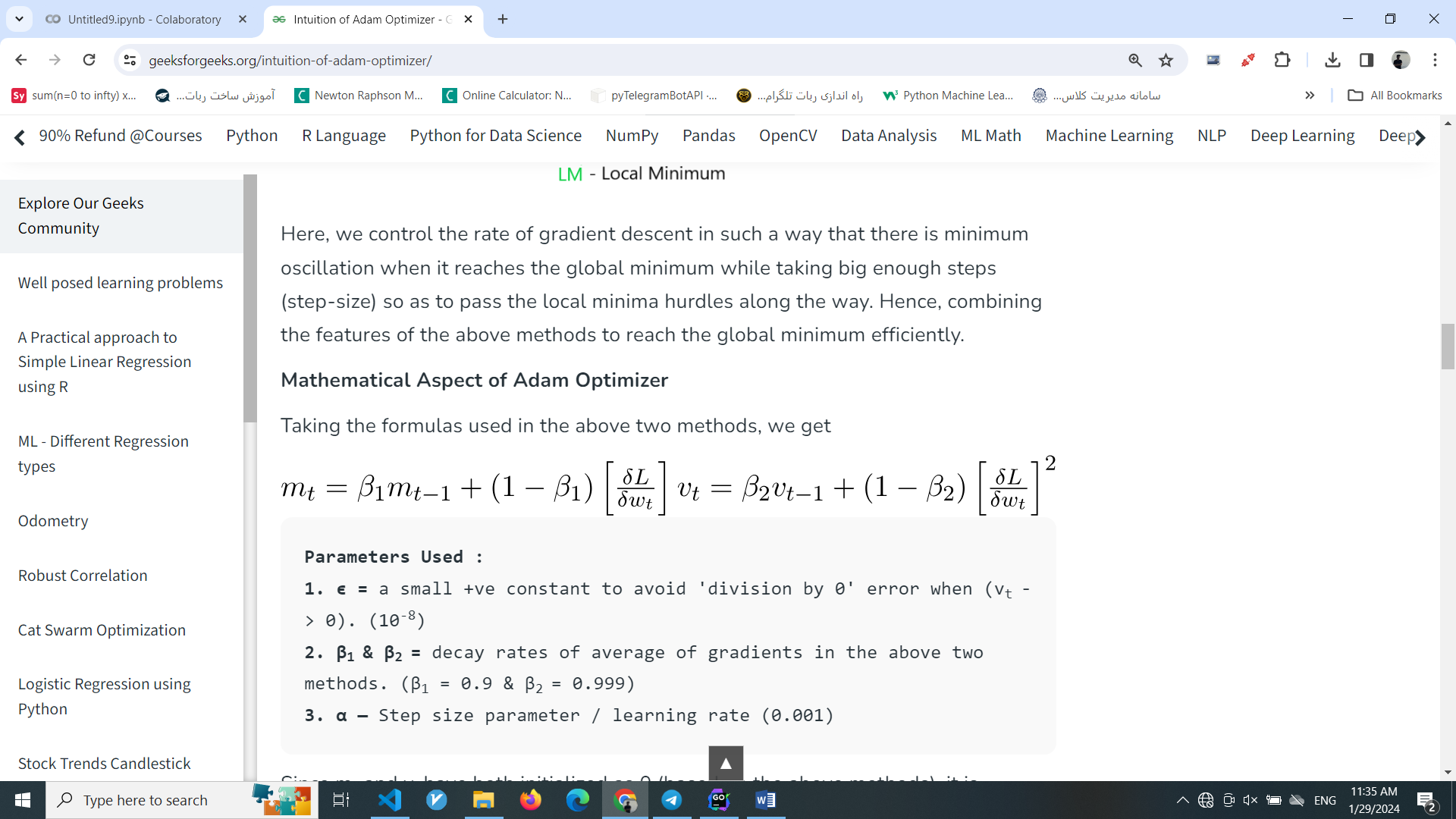
گزارش۱:

Linear یک لایه خطی از پرسپترون ها است که تعدادی ورودی و خروجی دارد. ReLU یک تابع فعال ساز یا activation function است که مقادیر منفی را صفر میکند و مقادیر مثبت را نگه میدارد. دلیل استفاده از تابه ReLU اینست که مدل ما غیر خطی شود زیرا اگر از چند لایه خطی بدون تابع فعال ساز غیر خطی بین آنها استفاده کنیم مدل نهایی هم خطی شده و انگار از یک لایه استفاده کرده ایم.

گزارش۲:

برای آپدیت وزن ها از optimizer استفاده میکنیم که در هر مرحله بر اساس مقدار loss و با الگوریتم backpropagation تمام وزن ها را آپدیت میکند.

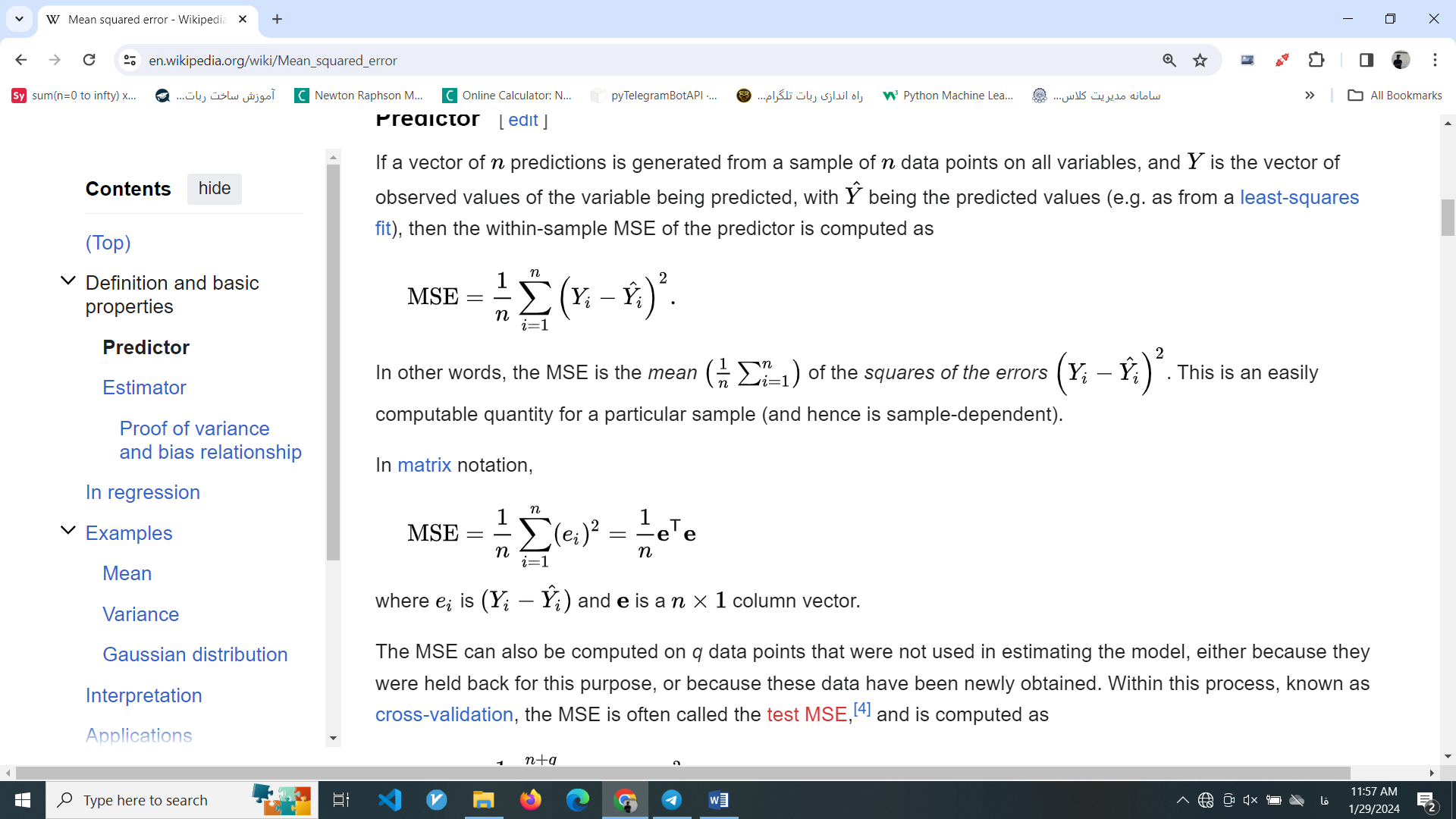
در الگوریتم adam از دو روش RMSProp وAdaGrad بهره گیری میشود برای آپدیت کردن وزن ها. به طور کلی این الگوریتم یک الگوریتم تصادفی یا stochastic است. فرمول محاسبه آن و همچنین فرمول آپدیت وزن ها در زیر آمده است که دارای دو هایپر پارامتر b1 وb2 میباشد که بر اساس آنها mt و vt حساب شد و سپس وزن ها آپدیت میشوند.



گزارش۳:

ابتدا باید خطا را بدست آوریم و سپس بر اساس خطا و به وسیله backpropagation وزن ها را آپدیت کنیم. در واقع بدون loss ما اطلاع نداریم که وزن های ما به چه اندازه خوب یا بد بوده که آنرا خوب در نظر بگیریم یا بد و اگر بد است چگونه آنرا آپدیت کنیم.

Mse یا mean squared error یک تابع محاسبه خطا است که میانگین مربعات اختلاف مقدار واقعی و مقدار پیش بینی شده توسط یک مدل روی تعدادی داده را حساب میکند. فرمول آن برای n داده به صورت زیر میشود.



گزارش ۴:

این کار برای ایجاد تعادل بین eplorationو expansion است. با اینکار بیشتر اوقات طبق بهترین استراتژی پییش میرویم و گاهی اوقات حرکت های تصادفی انجام میدهیم تا محیط را بیشتر بشناسیم و به استیت های مشاهده نشده که ممکن است بهتر از استیت های مشاهده شده باشند میرویم. با اینکار استیت های بیشتری میبینیم وآموزش بهتر میشود.

با زیاد کردن epsilon تعداد حرکت های تصادفی بیشتر میشود و محیط را بیشتر میشناسیم ولی دیرتر همگرا میشویم و دیرتر استیت های بهتر و حرکات بهتر را تشخیص میدهیم. با کم کردن آن حرکات تصادفی کمتر شده و زودتر همگرا میشود ولی ممکن است در مینیمم محلی گیر کند و استیت های بهتر را پیدا نکند.

گزارش ۵:

بر اساس استیتی که پس از انجام یک حرکت به دست می آید و ریوارد های بدست آمده مقدار q جدید را برای استیت جدید به دست میاوریم که در q\_targets بدست می آید. روش محاسبه آن هم به اینگونه است که qبدست آمده برای استیت قبلی را با q استیتی که پس از انجام اکشنش انتخابی انجام میشود بدست میاید را با یک ضریب تخفیف جمع میکنیم که ضریب تخفیف در اینجا گاما است. دلیل این فرمول اینست که باید مقدار ریواردی که تا الان بدست آوردیم را با ریواردی که در آینده میتوانیم بدست بیاوریم در نظر بگیریم. یک ضریب تخفیف هم در نظر میگیریم تا برای زودتر رسیدن ریوارد هم اولویت در نظر داشته باشیم.